

```

1  GCCAGCTGGG GTTACTTTAA AAAACATGCT CCATGTGCAT CCCTCTTGAA
51  GCTTCGCACT CTGTTGAAGA GGACACTCAT CCCAGTCATT ATTTAGAAGC
101 AAGGTCCTTG AATGAGCGAG ATTATCGGGA CCGGAGATAC GTTGACGAAT
151 ACAGGAATGA CTAATGTGAA GGATATGTTC CTAGACATTA TCACAGAGAC
201 ATTGAAAGCG GGTATCGAAT CCACTGCACT AAATCTTCAG TCCGCAGCAG
251 GAGAAGCAGT CCTAAAAGGA AGCGCAATAG ACACTGTTCA AGTCATCAGT
301 CACGTTGCGA GAGCCACCGA AGGAAAAGAT CCAGGAGTAT AGAGGATGAT
351 GAGGAGGGTC ACCTGATCTG TCAAAGTGGA GACGTTCTAA GAGCAAGATA
401 TGAAATCGTG GACACTTTGG GTGAAGGAGC CTTTGGCAAA GTTGTAGAGT
451 GCATTGATCA TGGCATGGAT GGCATGCATG TAGCAGTGAA AATCGTAAAA
501 AATGTAGGCC GTTACCGTGA AGCAGCTCGT TCAGAAATCC AAGTATTAGA
551 GCACTTAAAT AGTACTGATC CCAATAGTGT CTTCCGATGT GTCCAGATGC
601 TAGAATGGTT TGATCATCAT GGTCATGTTT GTATTGTGTT TGAAGTACTG
651 GGAAGTAGTA CTTACGATTT CATTAAAGAA AACAGCTTTC TGCCATTTCA
701 AATTGACCAC ATCAGGCAGA TGGCGTATCA GATCTGCCAG TCAATAAAAT
751 TTTTACATCA TAATAAATTA ACCCATACAG ATCTGAAGCC TGAAAAATTT
801 TTGTTTGTGA AGTCTGACTA TGTAGTCAAA TATAATTCTA AAATGAAAACG
851 TGATGAACGC ACACTGAAAA ACACAGATAT CAAAGTTGTT GACTTTGGAA
901 GTGCAACGTA TGATGATGAA CATCACAGTA CTTTGGTGTG TACCCGGCAC
951 TACAGAGCTC CCGAGGTCAT TTTGGCTTTA GGTTGGTCTC AGCCTTGTGA
1001 TGTTTGGAGC ATAGGTTGCA TTCTTATTGA ATATTACCTT GGTTTCACAG
1051 TCTTTCAGAC TCATGATAGT AAAGAGCACC TGGCAATGAT GGAACGAATA
1101 TTAGGACCCA TACCACAACA CATGATTCAG AAAACAAGAA AACGCAAGTA
1151 TTTTCACCAT AACCAGCTAG ATTGGGATGA ACACAGTTCT GCTGGTAGAT
1201 ATGTTAGGAG ACGCTGCAAA CCGTTGAAGG AATTTATGCT TTGTCATGAT
1251 GAAGAACATG AGAAACTGTT TGACCTGGTT CGAAGAATGT TAGAATATGA
1301 TCCAAC TCA AGAATTACCT TGGATGAAGC ATTGCAGCAT CCTTCTTTG
1351 ACTTATTA AAAAGAAATGA AATGGGAATC AGTGGTCTTA CTATATACTT
1401 CTCTAGAAGA GATTACTTAA GACTGTGTCA GTCAACTAAA CATTCTAATA
1451 TTTTGTAAA CATTAAATTA TTTTGTACAG TTAAGTGTAA ATATTGTATG
1501 TTTTGTATCA ATAGCATAAT TAACTTGTTA AGCAAGTATG GTCTTGATAA
1551 TGCATTAGAA AAATTA AAAAT TAATTTTTCT TTTTGAAATT ACCATTTTTA
1601 AATACCTTTG AAATATCCTT TGTGTCCAGT GATAAATGTG ATTGATCTTG
1651 CCTTTTGTAC ATGGAGGTCA CCTCTGAAGT GATTTTTTTT GAGTAAAAGG
1701 AAATCTTGAG TACTTTATAT TCTTAAAGGA ATATTCTTTA TATACTTCAA
1751 ATTTAGAACT TAACTTTAAA AGTTTTTCTT CTGTAATTGT TGAACGGGTG
1801 ATTATTATTA ACTCTAGATA AGCAGGTACT AGAAACCAAA ACTCAGAAAA
1851 TGTTTACTGT TAGAATTCTA TTAAATTTTA AGTGTGTAT TCTTTTTCAT
1901 TGGGTGATGT CAGGGTGATA ACCAGACATT CATGGAAAGG CATGCAGTTT
1951 GTCCATTGTG ACAGTTTGTG TAATAAAACC ACATACACAC TTTATTTAAG
2001 ATTA AAAATCT AACTGGAAAG TCAGCTTGGA AAATGGACAT TTCCAAGTAT
2051 GTTTGGTGAG TCACAGATAT AAAAATAGAA ATTCTGATGA GAGGTTTCAG
2101 TTTTAAATAC CAAGTCCTTA GGAGTCTTAA CATTGGCCAG CATCTGTTTA
2151 TCAAATGACA TAAATACGTA AACCTATAAG AATTAAGTTT ATTAATTAGG
2201 CAATTTATGT CTGTGATAAT TCTTACGGGA GAAAGAGGAT TTGATTGGAA
2251 AGCAGTTTGG GAAGAAAAGT CTGCTGAAAT TTCCAGAATT TAATTGATTG
2301 GTTACATAAA CTTTTTGACT TCAGAAAAAA AAAATAAAAA AACAAAAAAA
2351 AAAC

```

```

(SEQ ID NO: 1)

FEATURES:
5'UTR: 1 - 32
Start Codon: 33
Stop Codon: 1368
3'UTR: 1371

```

FIGURE 1A

Homologous proteins:

Top 10 blast hits:

Sequences producing significant alignments:		Score (bits)	E value
CRA 150000079514205	/altid=gi 10190706 /def=ref NP_065717.1 pr...	904	0.0
CRA 18000005115066	/altid=gi 6671766 /def=ref NP_031740.1 CDC ...	883	0.0
CRA 335001098680506	/altid=gi 11416272 /def=ref XP_003664.1 si...	745	0.0
CRA 335001098687191	/altid=gi 11429914 /def=ref XP_002520.1 CD...	740	0.0
CRA 18000004973971	/altid=gi 4758008 /def=ref NP_004062.1 CDC-...	738	0.0
CRA 18000004935844	/altid=gi 110864 /def=pir A39676 protein ki...	718	0.0
CRA 18000004938713	/altid=gi 125290 /def=sp P22518 CLK1_MOUSE P...	716	0.0
CRA 114000015334919	/altid=gi 9437515 /def=gb AAF87326.1 AF2122...	700	0.0
CRA 18000004896888	/altid=gi 107458 /def=pir A38643 protein ki...	670	0.0
CRA 98000043608390	/altid=gi 12805489 /def=gb AAH02220.1 AAH022...	630	e-179

EST:

Sequences producing significant alignments:		Score (bits)	E value
gi 12603368	/dataset=dbest /taxon=96...	785	0.0
gi 2555404	/dataset=dbest /taxon=9606 ...	712	0.0
gi 10341364	/dataset=dbest /taxon=960...	549	e-154
gi 3733981	/dataset=dbest /taxon=9606 ...	450	e-124
gi 900131	/dataset=dbest /taxon=9606 /...	432	e-118
gi 6034370	/dataset=dbest /taxon=9606 ...	424	e-116
gi 2824947	/dataset=dbest /taxon=9606 ...	396	e-108
gi 7318123	/dataset=dbest /taxon=9606...	381	e-103
gi 10913732	/dataset=dbest /taxon=96...	335	2e-89

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|12603368 Bone osteosarcoma cell line
 gi|2555404 Breast
 gi|10341364 Uterus leiomyosarcoma
 gi|3733981 Fetal heart
 gi|900131 Infant brain
 gi|6034370 Colon-juvenile granulose tumor
 gi|2824947 Mixed
 gi|7318123 Colon-moderately differentiated adenocarcinoma
 gi|10913732 Bone marrow hematopoietic stem cells
 gi: 2824947 Pooled human melanocyte, fetal heart, and pregnant uterus
 gi: 10088906 nervous_normal
 gi: 9093801 leukopheresis myeloid cell

Tissue expression:

Leukocyte

FIGURE 1B

```

1 MCIPLEASHS VEEDTHPSHY LEARSLNERD YRDRRYVDEY RNDYCEGYVP
51 RHYHRDIESG YRIHCSKSSV RSRRSSPKRK RNRHCSSHQ SRSKSHRRKRS
101 RSIEDDEEGH LICQSGDVL RARYEIVDTLG EGAFGKVVEC IDHGMDGMHV
151 AVKIVKNVGR YREAAARSEIQ VLEHLNSTDP NSVFRCVQML EWFDDHGHVC
201 IVFELLGLST YDFIKENSFL PFQIDHIRQM AYQICQSINF LHHNKLTHTD
251 LKPENILFVK SDYVVKYNSK MKRDERTLKN TDIKVVDVFGS ATYDDEHHST
301 LVSTRHYRAP EVILALGWSQ PCDVWSIGCI LIEYYLGFTV FQTHDSKEHL
351 AMMERILGPI PQHMIQKTRK RKYFHNNQLD WDEHSSAGRY VRRRCKPLKE
401 FMLCHDEEHE KLFDLVRRML EYDPTQRITL DEALQHPFFD LLKKK
(SEQ ID NO: 2)

```

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

176-179 NSTD

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
CAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1	73-76	RRSS
2	97-100	RKRS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 8

1	69-71	SVR
2	72-74	SRR
3	76-78	SPK
4	94-96	SHR
5	277-279	TLK
6	303-305	STR
7	368-370	TRK
8	425-427	TQR

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 8

1	10-13	SVEE
2	25-28	SLNE
3	102-105	SIED
4	128-131	TLGE
5	209-212	STYD
6	247-250	THTD
7	292-295	TYDD
8	429-432	TLDE

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

Number of matches: 3

1	24-31	RSLNERDY
2	29-36	RDYRDRRY
3	55-61	RDIESGY

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

147-152 GMHVAV

FIGURE 2A

[7] PDOC00100 PS00107 PROTEIN_KINASE_ATP
Protein kinases ATP-binding region signature

129-153 LGEGAFGKVVVECIDHGMDGMHVAVK

[8] PDOC00100 PS00108 PROTEIN_KINASE_ST
Serine/Threonine protein kinases active-site signature

246-258 LTHTDLKPENILF

Membrane spanning structure and domains:

ndidate membrane-spanning segments:

Helix	Begin	End	Score	Certainty
1	324	344	1.141	Certain

BLAST Alignment to Top Hit:

>CRA|150000079514205 /altid=gi|10190706 /def=ref|NP_065717.1| protein
serine threonine kinase Clk4 [Homo sapiens] /org=Homo
sapiens /taxon=9606 /dataset=nraa /length=481
Length = 481

Score = 904 bits (2312), Expect = 0.0
Identities = 427/427 (100%), Positives = 427/427 (100%)
Frame = +3

Query: 87	HYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYHRDIESGYRIHCSKSSVRSRRSSPK	266
Sbjct: 55	HYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYHRDIESGYRIHCSKSSVRSRRSSPK	114
Query: 267	RKRNHRHCSSHQSRKSHRRKRSRSIEDDEEGHLICQSGDVLRLARYEIVDTLGEAFGKVV	446
Sbjct: 115	RKRNHRHCSSHQSRKSHRRKRSRSIEDDEEGHLICQSGDVLRLARYEIVDTLGEAFGKVV	174
Query: 447	ECIDHGMDGMHVAVKIVKNVGRYREAAARSEIQVLEHLNSTDPNSVFRCVQMLEWFDHHGH	626
Sbjct: 175	ECIDHGMDGMHVAVKIVKNVGRYREAAARSEIQVLEHLNSTDPNSVFRCVQMLEWFDHHGH	234
Query: 627	VCIVFELLGLSTYDFIKENSFLPFQIDHIRQMAYQICQSINFLHHNKLTHTDLKPENILF	806
Sbjct: 235	VCIVFELLGLSTYDFIKENSFLPFQIDHIRQMAYQICQSINFLHHNKLTHTDLKPENILF	294
Query: 807	VKSDYVVKYNSKMKRDERTLKNTDIKVVDGFSATYDDEHHSTLVSTRHYRAPEVILALGW	986
Sbjct: 295	VKSDYVVKYNSKMKRDERTLKNTDIKVVDGFSATYDDEHHSTLVSTRHYRAPEVILALGW	354
Query: 987	SQPCDVWSIGCILIEYYLGFTVFQTHDSKEHLAMMERILGPIQHMIQKTRKRKYFHHNQ	1166
Sbjct: 355	SQPCDVWSIGCILIEYYLGFTVFQTHDSKEHLAMMERILGPIQHMIQKTRKRKYFHHNQ	414
Query: 1167	LDWDEHSSAGRYVRRRCKPLKEFMLCHDEEHEKLFDLVRRMLEYDPTQRITLDEALQHPF	1346
Sbjct: 415	LDWDEHSSAGRYVRRRCKPLKEFMLCHDEEHEKLFDLVRRMLEYDPTQRITLDEALQHPF	474
Query: 1347	FDLLKKK	1367
Sbjct: 475	FDLLKKK	481 (SEQ ID NO:4)

FIGURE 2B

```

>CRA|18000004973971 /altid=gi|4758008 /def=ref|NP_004062.1| CDC-like
    kinase1; CDC-like kinase 1 [Homo sapiens] /org=Homo
    sapiens /taxon=9606 /dataset=nraa /length=484
    Length = 484

Score = 738 bits (1884), Expect = 0.0
Identities = 352/429 (82%), Positives = 382/429 (88%), Gaps = 2/429 (0%)
Frame = +3

Query: 84  SHYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYHRDIESGYRIHCSKSSVRSRRSSP 263
          SHYLE+RS+NE+DY  RRY+DEYRNDY +G  P H  RD ES Y+ H SKSS RS RSS
Sbjct: 54  SHYLESRSINEKDYHSRRYIDEYRNDYTQCEPGRQRDHESRYQNHSSKSSGRSGRSSY 113

Query: 264  KRK-RNRHCSSHQ-SRSKSHRRKRSRSIEDDEEGHLICQSGDVLRARYEIVDTLGEGAFG 437
          K K R  H +SH+ S  KSHRRKR+RS+EDDEEGHLICQSGDVL ARYEIVDTLGEGAFG
Sbjct: 114  KSKHRIHHSTSHRRSHGKSHRRKRTRSVEDDEEGHLICQSGDVLRARYEIVDTLGEGAFG 173

Query: 438  KVVECIDHGMGMHVAVKIVKNVGRYREAAARSEIQVLEHLNSTDPNSVFRCVQMLEWFDH 617
          KVVECIDH  G HVAVKIVKNV RY EAARSEIQVLEHLN+TDPNS FRCVQMLEWF+H
Sbjct: 174  KVVECIDHKAGGRHVAVKIVKNVDRYCEAARSEIQVLEHLNTTDPNSTFRCVQMLEWFEH 233

Query: 618  HGHVCIVFELLGLSTYDFIKENSFLPFQIDHIRQMAYQICQSINFLHHNKLTHTDLPEN 797
          HGH+CIVFELLGLSTYDFIKEN FLPF++DHIR+MAYQIC+S+NFLH NKLTHTDLPEN
Sbjct: 234  HGHICIVFELLGLSTYDFIKENGFLPFRLDHIRKMAYQICKSVNLFHSNKLTHTDLPEN 293

Query: 798  ILFVKSDYVVKYNSKMKRDERTLKNTDIKVDFGSATYDDEHHSTLVSTRHYRAPEVILA 977
          ILFV+SDY  YN K+KRDERTL N DIKVDFGSATYDDEHHSTLVSTRHYRAPEVILA
Sbjct: 294  ILFVQSDYTEAYNPKIKRDERTLINPDIKVDFGSATYDDEHHSTLVSTRHYRAPEVILA 353

Query: 978  LGWSQPCDVWSIGCILIEYYLGFTVFQTHDSKEHLAMMERILGPIQHMIQKTRKRKYFH 1157
          LGWSQPCDVWSIGCILIEYYLGFTVF THDSKEHLAMMERILGP+P+HMIQKTRKRKYFH
Sbjct: 354  LGWSQPCDVWSIGCILIEYYLGFTVFPTHDSKEHLAMMERILGPLPKHMIQKTRKRKYFH 413

Query: 1158 HNQLDWDEHSSAGRYVRRRCKPLKEFMLCHDEEHEKLFDLVRRMLEYDPTQRITLDEALQ 1337
          H++LDWDEHSSAGRYV R CKPLKEFML  D EHE+LFDL+++MLEYDP +RITL EAL+
Sbjct: 414  HDRLDWDEHSSAGRYVSRACKPLKEFMLSQDVEHERLFDLIQKMLEYDPAKRITLREALK 473

Query: 1338 HPFFDLLKK 1364
          HPFFDLLKK
Sbjct: 474  HPFFDLLKK 482 (SEQ ID NO:5)

```

Hmmer search results (Pfam):
 Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	272.4	5.9e-78	1
CE00022	CE00022 MAGUK_subfamily_d	26.7	8.6e-08	2
CE00204	CE00204 FIBROBLAST_GROWTH_RECEPTOR	3.4	2.3	1
PF00548	3C cysteine protease (picornain 3C)	1.6	7.7	1
CE00031	CE00031 VEGFR	0.7	2.5	1
CE00289	CE00289 PTK_PDGF_receptor	-49.9	0.0045	1
CE00292	CE00292 PTK_membrane_span	-102.3	0.0063	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-117.7	0.97	1
CE00291	CE00291 PTK_fgf_receptor	-138.4	0.73	1
CE00290	CE00290 PTK_Trk_family	-173.0	0.0023	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-239.0	0.0019	1
CE00288	CE00288 PTK_Insulin_receptor	-240.3	2.7	1

FIGURE 2C

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00204	1/1	128	138 ..	515	525 ..	3.4	2.3
CE00031	1/1	120	175 ..	873	934 ..	0.7	2.5
CE00289	1/1	120	223 ..	1	109 []	-49.9	0.0045
CE00022	1/2	306	331 ..	191	216 ..	4.9	0.23
CE00288	1/1	125	353 ..	1	269 []	-240.3	2.7
CE00291	1/1	123	368 ..	1	285 []	-138.4	0.73
PF00548	1/1	370	378 ..	175	183 .]	1.6	7.7
CE00287	1/1	123	379 ..	1	260 []	-117.7	0.97
CE00290	1/1	124	379 ..	1	282 []	-173.0	0.0023
CE00292	1/1	123	381 ..	1	288 []	-102.3	0.0063
CE00022	2/2	414	437 ..	258	281 ..	21.6	2.8e-06
PF00069	1/1	123	439 ..	1	278 []	272.4	5.9e-78
CE00016	1/1	66	445 .]	1	433 []	-239.0	0.0019

FIGURE 2D

1	GCAGAAAAGT	ATAAAGATGG	TAATCTCTGT	AGGAAATTAG	TCCCCATTAT
51	TTAGCTGTAA	AATTATAATT	AAAAAAAAAA	ATCTTTGTTT	CTAAATCTTT
101	GCCACTGATT	ATTTCTTGAA	AATACACTCC	AGGAAGAAGC	ATTTTTAAGT
151	TAAAGCATGT	GAACCTTTAT	TTCTTGCTAC	AGGTTTCATAT	TTCTTTTTCT
201	AGAGAGTTTG	CCAAATTATA	CAACGTGCTC	CTTCATGCTC	TCACCAATCT
251	TGGCTGTTTT	GAAAGGCCAA	GAATAATGTT	TTGATTAAAC	TGAATTTTTA
301	AATTTCTAAC	GAATTTGTCC	GCTGTCATAT	ATTTATTGAT	CATTTGAACA
351	TCTTTTTATT	CTTAGCCTAT	TTATTAAAGT	ATTTTTATTG	ATTTAGAAGA
401	GCTTTTTATT	ACAATATTTT	AACCATTGT	CATATATATA	TTGCATAGTG
451	TCTTTTCTTT	ATGATTTGTC	TTTTGGAGGT	AGCCTGTGAA	TTGGTCTCCC
501	TTTCTACAGG	CTTAGTTAAT	CCATTCTGCA	TTAGAAAGAC	TGATGTGGCT
551	GTA AACCTA	CCTTTATATA	TTGTGGTCAG	AAGCCTGTAA	CATAAAGTAT
601	CAAGTCTTAA	ACCAGTGATT	CTCCAACTTT	AGTGTGAATA	AGAATCACCT
651	TGGAGGTATG	CTGACCAGAT	TTACAGTCAG	TGAGTATGAC	CTAAGGCCCA
701	GGGTACCATT	TTTTAATAAG	AACTCCATAT	TTGATACTGT	TGATAAATAG
751	ACCGTCCTTT	GAGAAATAAT	ACTCTTTAGC	CTAGCACGCA	GGGTTTTTAA
801	TGATGCTATT	CTCAGCTTAC	TTATTTGTCT	ACATTCCCCT	ATGTGAAAAT
851	TGCTCTTGCT	GGGATTGTCT	TTTTCTGAG	TAATGCATAG	ACAATTCCAT
901	CTCTAAGCCA	TTGTGGCTAA	AAGTGCCATA	TGAATTTAAG	ATGGTAATAT
951	GCCATTCTTC	TCCCCCGGAA	TTTCTTCTGT	ATTCTACTTT	TTCCAAATCC
1001	TGGCTTCCCT	TTAAGATGCA	ACTCTATTTT	CATCTTTTTT	GTAATTATTC
1051	TCTGACCATT	TTAAACAGAT	TTTTTCCCCC	ATCTCTGACT	CTAAGCACTC
1101	ATGTGTTGTA	ACCTTTTAGA	ATTTCTTACA	TTGTTGGATT	TTGTTTCATT
1151	TTTATGTGAG	TAATCTCAAA	TTGTTTCATT	TTTGTGGGCA	GGGACTTTGC
1201	CTTATATAAT	TTTTTTTTTA	TCTCCCACAG	GACCTGTGTG	GATATAAAAA
1251	CGAATGCCCT	TACCCTCATC	CGTCTTGGCT	ATTTGAAAGG	CTATAGTGAA
1301	ATATTCACTG	GGCATTCACT	GGATATTTTA	AAAAATTAAA	TCAGTCTGTT
1351	CATCCTGTCC	ATAGCCTGTG	TAATTCTGTA	GACTTTGTTT	ATATAATCTC
1401	TCAGCCTTGG	TCATTGGCCA	TTATCTATTG	AAGAGACTCT	CATCCTTTTA
1451	GTTTGTCTCT	ATGGTGTTC	CTCCCATGTT	TTGTTACTCT	ATACGTTGTT
1501	TATGGCTTAG	CAGCTCTAAT	TCCATGCAGT	ATTCCAGCTA	AAGATTGTTA
1551	GTGCTAGTTT	TTTCTAATAG	AAGGATTTTG	GACTTTTATG	GGAAGGATGC
1601	CCTTAAGAGT	ATGGTCACGT	CTAGCTTATT	GTATTGGTGA	TCTCTCCCTG
1651	ACGATTTCAA	GCCAACTGAT	CAGATCTCTA	ACCTAGACTA	CCCACAGTCT
1701	TACCCAAATA	TCCTGAGTTG	TTTCTCCAAT	AAAATACAAC	TTAAAGCTGA
1751	TGCTAGGGAA	AGAGAACCGG	GTTTCTGTAT	CTCCCCAGCC	TGGATTTGAT
1801	GCTAGCCCTA	TTGGGTAGTA	GTTGTAAAGA	TGCTTCTATT	TCTGCCTAAA
1851	CCAGCCCCCT	GGGAAAAAGA	ATGACAGCAT	ATTCTGGGGA	AAGGAAAGGG
1901	GTTGGTGAGG	GCAATCTAGT	CAACATCCGT	CACTCCATTG	CTTGTTAGGC
1951	TTATTTTAGC	CGATGTGTCT	GACTGGGCAG	GTGTCCCCTC	TCTCCCTCAG
2001	TGCTCCATGT	GCATCCCTCT	TGAAGCTTCG	CACTCTGTTG	AAGAGGACAC
2051	TCATCCCAGG	TAGAGAGGGG	GACGGGAAAC	TGGGCCAATT	GAATCTATGT
2101	CCTTTTCTTT	CCATCAGATC	AAGGCCACTT	AACTGGGATC	CATTGACATC
2151	CTGAGGCCCA	TGACCTTTGA	AATTCCTTGC	CAAGTTTTGT	TTATGTGTTT
2201	CTTAGGAAAG	AGAGTCCATG	GCTTTTCAAG	GATTTTCAAA	GGGATCTCTA
2251	GATTAAGACA	CGATGGCACT	AGATGATGGT	GTTTTCTGTT	GTTTCTTAGG
2301	TATTTCTCAA	ACAGGAATGA	CAGGAAATTA	GAAATGCAAA	GGGAAGTAGG
2351	GTGGTGGAAC	TATTGTAATG	CTAAACTACA	GGATCCCCTT	CTTATTTTAG
2401	GGGGATATAT	TTTAGATGCC	TTTGGCACAT	GAGGCAGTCC	TCAAAAGCTA
2451	TGTTTTCTAT	TTCTCAAAAC	GGAATAACAA	GGCTAGAAAT	GCAAAGAGTA
2501	GAGGAGACAT	GATAGATGCT	GTGTGTAATA	AAATTGGCCT	GTATAATAGT
2551	GGTTTGAAAA	TATTTTAGTT	TTTGTCACTA	ATGTTGTTAT	ACAACCTTGG
2601	TAAATCATTT	TTCTTCTAGG	GATCTTAATG	TAGTCGTCGG	TAAAATGAAA
2651	GGGCTGGAAT	ACATTTAAGG	CTCCTTATAG	CTCTAATATA	CCTTTCATGA
2701	AGGAATTCTC	TCTGTGCCAG	GGATATCTAA	AATGCTCTTA	CATTACAAGA
2751	GAAAGGAATC	CTTTTTGCCT	GCCTCTGATT	GTACCTCTGT	GAGAGACTAA
2801	GACAGCTTAG	ATACAGGTGC	AGAAGGTAAA	GGAACACTTA	ATCAAGTAAA
2851	CAC TAGACAT	GAATTAATGA	TTTGACTCAA	GCTTTATTCC	TTGGTGTGAA
2901	GTGCTTGACA	GCAAACTCTA	TAATGGGCCC	ATTTGCTTGT	TTGTTAAAGT
2951	AAAATTATTT	CTTAAGCTTT	ATGAGATAAA	TATAAATGCT	AATTCATCTG
3001	TTTGAATTTT	TTTCTTATAT	TGAGTTAGCT	GTTTAAGAAT	TTCTGAGAAA
3051	ATGTTTTGTT	TGAACCACAT	TATTGCAGAA	TGAAGAGAAT	AATTTGAAAT
3101	CTTTTAATGC	TTTTCAGTGC	ATTATTTAGA	AGCAAGGTCC	TTGAATGAGC
3151	GAGATTATCG	GGACCGGAGA	TACGTTGACG	AATACAGGAA	TGACTACTGT

FIGURE 3A

3201	GAAGGATATG	TTCTAGACA	TTATCACAGA	GACATTGAAA	GCGGGTATCG
3251	AATCCACTGC	AGTAAATCTT	CAGTCCGCAG	CAGGAGAAGC	AGTCCTAAAA
3301	GGAAGCGCAA	TAGACACTGT	TCAAGTCATC	AGTCACGTTT	GGTATGATTG
3351	GTTTGTGTTT	CAATTTGAGT	GGAGTTTTAT	TTGTGTGTAC	TCTTAACGAG
3401	CTGATAAGTT	TCTAATTTTT	TATATATATA	TATATAAAAT	ACTATTTGGA
3451	TATATTATAA	TTGTATTTAT	ATTACTTAAA	TCCTTAAAGG	AAACCTCCAA
3501	ATTCTTGTAG	CTGATCTGTA	TATTTATTAG	CTAGCCCTCA	TTTGCCCA
3551	TTTCCTCATA	TTCTGCAGAC	CAGATAATGA	GTTTATTGAT	TTTAATAATA
3601	AAACTATTTT	TTTATTTGTA	ACATATTCTT	ATGAAAAAAT	CATGCACCCA
3651	TATCTTTTCT	TTTATCTTAA	GCATTTTTTT	TTTCTTAGAA	ACCCTTTATC
3701	TGGTACTTGA	AAATAAATGT	GAAATATTGC	ACTGGTGGAC	ACCTGAATGT
3751	TACTAACCTG	CATAGAGCAT	AGTTCCATAG	TCCAGTGCAT	CATTGTCTGC
3801	AATGAATTCT	TTTGAAGTTG	TGAAAAATGGG	TGCTGAATGG	GAAACATCCA
3851	AAAAGTCTGC	CCCCCCTTTT	TTTTTTTTTAA	CACTCAGACA	TCTTCACCTG
3901	CTTGAACAGT	GAACTTTGAA	TTAGTTTCTC	CCCAAGTTTT	CTTCAGTAAA
3951	ACTAGTTTTT	ATTAGATTGA	ACATTGAAAT	TAACTAGCCT	TTATTTTCCC
4001	CTTTTATTTT	AATCATGTAT	ATTTTAAAAAT	ATTGCTAAAT	TAGAATAATT
4051	TCAAATAGTC	TTGACATTTT	AAAAACATTTT	TCTGAAAAAC	TAGACATCTC
4101	AATTCACAGC	ATATGCTGTT	TATAGCAAGA	GATAAGTAAA	TCATGACATT
4151	GCATTCTTTA	AATTTTCAGAC	TTCAATTAAA	TCAGTATTTT	AAAGAGACAA
4201	TTGTGTTGTT	TTTTTCTATT	GCCACTTTAA	GTATCTTATC	TGAAAAATCTG
4251	TTCTTTGCCA	TGTTTTTCTT	CTGTAACATA	AACTGTGCCC	TGTGAATTTT
4301	TGGGGACTGA	ATTTGAAATT	AGCTCTGCCA	ACTGTTCTGT	GCCTGGTGTCT
4351	TATCTGAATG	CCTGAATATC	TCCCCGCTGA	ATGAATTGCG	TATTCTGCCC
4401	TGAATTCAT	CTGATATATT	GATTGGCTGG	ACGATCTTGG	TGCTGCCCAC
4451	TTGCCGTTCC	AGAAGAGCCA	CCGAAGGAAA	AGATCCAGGA	GTATAGAGGA
4501	TGATGAGGAG	GGTCACCTGA	TCTGTCAAAG	TGGAGACGTT	CTAAGAGCAA
4551	GATGTATAGA	ATATTTTTCA	ACACTTTTTA	AACTTTGCAG	AAAGAATAAT
4601	CTTTTTAAGA	ATAGTTTGTC	AGCGGGGGGC	TAAAGAACTC	TTCATTGCTT
4651	TTTTATTTTG	CTTTTTGTGG	GTTTTGTTGT	TCTTTTATAT	TTCTTCTTTT
4701	CTGTAGAATT	TAAATATTTT	TATTCTAAAG	TTCCAAAATA	ATCAGTGGAA
4751	TTTGAGATTA	GAGCAAGAAA	GATAGCTCTA	TCTAATTGTT	TTTGTAGCAG
4801	CTGAAACTAA	AATAATTTGA	GTGCTGAAAC	CTTAGTTATG	CTTTGTTAGA
4851	GATCATTTGA	AAATATTCCA	CACTTAAGCA	TTTATTGTTT	GAAGAACTAG
4901	ACAGTTTGTA	CTCAGGTACT	TACACCTCTT	TTTCCCTCCT	CACTCTAGAT
4951	GAAATCGTGG	ACACTTTGGG	TGAAGGAGCC	TTTGGCAAAG	TTGTAGAGTG
5001	CATTGATCAT	GGCATGTAAG	TTTGTTTTTT	CCTTTTCAAA	CATTCTGATG
5051	TTTTTGGTGG	GGAAAGATTG	ATAATTCAGA	TGAAATTTTA	TTTTTTTATT
5101	TATTTGAGAT	AGGGCCTCTG	TTGCCCAGGC	TTGAGTGCAG	TGGTGCTATC
5151	TTGGCTCACT	GCAACTGCCG	CCTCCCGGCT	TCAAGTGATT	CTCCTGCTTC
5201	AGCCTCTCAA	GTAGCTGGGA	TTACAGGAGC	CTGCCACCAC	ACCTAGCTAG
5251	TTTTTGATT	TTTAATAGAG	ATGGGGTTTC	ACCGTGTTGG	CCTGGGTGGT
5301	CTCGAACTCC	TGACCTCAAG	TGATCTACCC	GCCTCAGTTT	CCCAAAACGT
5351	TGGGATTACA	AGCCTGAGCC	CCTGTGCCCG	GCCAAGATGG	AATATATTTT
5401	AAATGGTAGC	CACGTGTTTT	GGGGGGTAAA	TTACTCACCA	AAGTTTCTTG
5451	AACTTTGAT	GATTTATTTA	CCGTGAATGT	GGATCTTAAG	AATGCTGACT
5501	GCCGGGCACA	GTGGCTCACT	CCTGTAATCG	CAGCACTTTG	GGAGGCCAAG
5551	GCAGGTGGAT	CACCTGAGGT	TGGGAGTTCA	AGACTAGCCT	GACCAACATG
5601	GAGAAATACA	TTCTCTACTA	AAAAATACAAA	ATTAGCCAGG	TGTGGTGGCA
5651	CATGCCTGTA	ATCCCAGTTG	CTTGGGAGGC	TGAGGCAGGA	GAATCACTTG
5701	AACCCAGGAG	GGGAGGAAGG	CGGAGGTTGC	GGTGAGCCAA	GATTGTGCCA
5751	TTGCACTCCA	GCCTAGGCAA	CGAGTGAAAA	TCCGTCTCAA	AAAAAATAAA
5801	AATAAAAAAA	AAGAATGATG	ACAAATTTCA	ACAGGGGGAA	ATCATTGAAA
5851	TTAAAGTGGA	TGTTCAAGTG	AAGGAATTTT	CCAGAACTCC	AGAACTGAGG
5901	CCCTTGACCC	TGTATATAAG	ATTTGGCAAT	TTCGGATTAC	AGAGGCAATA
5951	AAGCATGTCT	AATCTTAAAT	GTTAAGAGTT	AGCTTCCTAA	ACTATAAAGA
6001	CATTTTATTA	TCTAGGGCCT	AGAGAATAAA	GTTTGTGATT	TGACCCTTTC
6051	TGCCTCATTT	TACCGTTTTT	CTCTAGGACC	TCTATTTTGT	GGCTTGAAAA
6101	CTTTTGTAAG	AGAAGCTCTT	AGAACTTTTG	CGAAACTTCA	CATTTCTAAA
6151	ATGACAAAAT	TTTTTATCAT	AAATTATTTG	GGAAGGATGT	AATTTCCAAC
6201	CTGTTGTAAG	TATTAATATT	AAAAAATAAA	ACTTACCTCT	CTCTAAATGC
6251	ATTTTCAGGGA	ATCTAAATAC	CATAGCAGCT	TGATACCTAC	CATCATCCAT
6301	AAACAACTC	TTCTTGAATA	CTTAGAAATG	TTTTATTATT	GAATTTATTG
6351	TCATTTCACT	TTCCATAAAT	ACTATCCTAA	ATTATCCCCA	CATTTTGCTT
6401	TTCTGCAACA	AATATGTGAA	TGTAAATTGA	ACTTTAAAGT	ATTTTGAAAT

FIGURE 3B

6451	ATTTTCAGAC	TTACAGAAAA	ATTGATAAAA	TAGTTCAAAG	AATTCCCATA
6501	TATTCCAAAT	GTTAACCTAT	TTTCCAAATG	TTTACATTTT	ATAAGATTTG
6551	CTTTATCATT	ATACATACAT	TTGTTTTCAA	ATTTTGCCAA	CTAATCTGCA
6601	GACTTTATTC	AGATTTTACC	AGTCATCCCA	TTAATGTCCT	TTTAGAATTT
6651	CTTGAAAGTC	TAAGTCTTGG	TGTATTTAAT	GAAATGTATC	TTAAAAACAA
6701	TTTTTTTTTA	ATGAGATGGA	GTCTCACTGT	GTTGCTCTGG	CTGGTGTGGA
6751	ACTCCTGGCC	TCAAGTGATC	CTTCTGCCTC	AGCCTCCCAT	AGTGCTGGGA
6801	TTACAGGGTG	TGAGCCCTGT	AGTCACGTGT	GGCACACACC	TGTACCACAT
6851	CTGGCCTGGA	ATGTTTTCTT	TATTGGGGCA	GTTGAGGCCT	CTAAAAAATG
6901	AGTACATATA	GCCATAGATA	AATATCTGAC	TGTCTAGCAT	TGTATGTTTT
6951	CTTTTTTCAT	TTTCGTGGAT	ACAAGCACTG	AGAAAACTTT	TTGGTCATAT
7001	AATTAAATAG	ATAGGAGTAG	AAGCTTTGTC	ACAGTAATCT	TATTAGAGTT
7051	CTTTTAAGTC	TTGAGGTATA	TGCCAAGCAT	TAAAAAATTT	TTTTAGTGAC
7101	TTATCAGTTC	ACATTCGTTG	GGGCCTTGTT	GAAAGCAATG	AACTGGAAAC
7151	CACTGGATGT	GGAAAAAGGT	TTTGATATCCA	GCCATTAGAA	TACGTGTTTG
7201	TTTGCCCCAA	ATGTTTTTAT	AGCCTAGGGC	ATACATCCTG	TTACACTAGT
7251	AAGAGATGGG	TATGGTTTTG	TAAAGTGGAA	GGGTCATAGT	GAAAAAGAA
7301	GCTTGAATGC	TGGCTCATCT	GTAGGTAGAT	TAGGTTTAAA	AAGGAAGACA
7351	AAAATAAATT	GAAGATTTGC	AACATTTATG	GCTCTATACT	TTTTAGGAAG
7401	CATTCTTACA	GATGCCGCAG	TCTAAAGCCC	ACTGCCCTCC	CCTGTAGCTG
7451	TTTCTGTATA	CTGGCATCAG	TGCATCTGCT	AAGGTTTTTC	TGGGCTTCAT
7501	TACTTAGAGT	TGGGGTCTCC	TTTACCTGGA	TGTTTCCTTC	CCAATCTGAC
7551	AAACTCCCAG	CTATCTTTCA	GGACTCAGTT	CTGTGTCACC	TCTTCTGTGA
7601	AGAAGTCTAA	GTTGTTTCTG	TGTCTGTCTT	TTCCATTAGA	CTTTGAAGTA
7651	CGTAGGGACA	CACCCCGTCT	TTTAATCACT	AATATCTGTG	CATTGCCTGG
7701	CACAGAGTAG	GCCTAGCCTG	GTAATGAAT	GAATGCTTTC	AACAGTAGCA
7751	TATCCTATTT	TTGGTTTACA	TTTGATATA	TCTTTTAAAA	CTGTTGTTGT
7801	ATAAAATGTA	ATTAATTTTA	AAATTCTAGG	AGCAAACGTT	AAAACTCATA
7851	AGTATTAAGG	GAATTATCAC	TTCATATAAA	GTATTTTATC	AAAATGTTTT
7901	AAGAAGATGT	TATATGGAAT	CTGCTATAAT	ATGTTCTGAA	AGATTATTTT
7951	AAATGGCATA	GAGGAATTGG	TAATTAAGAT	TATGCTTTAG	AGCATAACAT
8001	GGCTTCAGCT	CACTCTTGTA	CATTTATCAT	TTTATCTTTA	ATTTTATTTT
8051	TAAGGGATGG	CATGCATGTA	GCAGTGAAAA	TCGTAAAAAA	TGTAGGCCGT
8101	TACCGTGAAG	CAGCTCGTTC	AGAAATCCAA	GTATTAGAGC	ACTTAAATAG
8151	TACTGATCCC	AATAGTGTCT	TGTAAGTATA	ACTTTCACCT	AGGAGCCATC
8201	ATATTACATG	AAATATTCAG	GTTTCCATAA	ACTGAATTAT	TATTTTGCTC
8251	TGTTTTAGCC	GATGTGTCCA	GATGCTAGAA	TGGTTTGATC	ATCATGGTCA
8301	TGTTTGTATT	GTGTTTGAAC	TACTGGGACT	TAGTACTTAC	GATTTTCATTA
8351	AAGAAAAACAG	CTTTCTGCCA	TTTCAAATTG	ACCACATCAG	GCAGATGGCG
8401	TATCAGATCT	GCCAGTCAAT	AAATTGTAAG	TACACTTGAT	AAATCTTTAT
8451	TTTTATTTAT	TTATTTATTT	ATTTATTTTG	AGACGGAGTC	TCGCTCTGTC
8501	ACCCAGGCTG	GAGTGCAGTG	GCGCTCTCGG	GTCCCAGCAA	GCTCAGCCTC
8551	CCGGGTTTAC	GCCATTTTCC	CGCCTCAGCC	TCCCGAGTAG	CTGGGACTAC
8601	AGGCGCCAC	CACCATGCCC	AGCTAATTTT	TTGTATTTTT	AGTAGAGATG
8651	GGATTTTACA	GTGTTAGCCA	GGATGGTCTC	GATCTCCTGA	CCTTGTTGATT
8701	GCCCCCCTCG	GCCTCCCAAA	GTGCTGGGGT	TATAGGCGTG	AGCCACTGTG
8751	CACAGCAATA	AATCTTTATT	TTTAAATATT	TTTTATGTTT	GTACCTCCTT
8801	AACAATTAAG	ATAAATCTTT	AAGCACCAGA	AAACTTGTTT	TTATTATACA
8851	AGCTATATAT	CCAAATGTTG	TCACTAAAAA	AACAGACATT	TTACAAGTAA
8901	AGATGAATCG	TCTCTTGACC	ACTATATCCT	TTGCCAGTCC	TCCTTTCCCT
8951	CCTAGTACAA	ATTAAGTTTG	TAAGTGAAAC	TAATAATGTG	CTTTTGTTCT
9001	CTTGATGTTT	TACATCATAA	TAAATTAAAC	CATACAGATC	TGAAGCCTGA
9051	AAATATTTTG	TTTGTGAAGT	CTGACTATGT	AGTCAAATAT	AATTCTAAAA
9101	TGGTAAGTTA	AAGACTTGTT	TTAATTTGGG	TGGTTGTCTT	TAAAATTAAT
9151	TTAACTTGAT	GATCTTTGGA	TGAGGAATTT	CACTTCTGAG	CCTTATTATA
9201	TCCTGTTGTT	TAACCAAAAA	GAAGTAATCC	TTCTTTGCCT	TTCTCATGAG
9251	CTTACTTTGA	CAATCAAGAA	GATAATTTCAT	GTGCTGGCCT	TTTGAGTAGC
9301	GCTATAAAAT	GTATCTATTG	AGTTTCATGT	TTACTCAACT	GTGTCTCTCT
9351	AGAAACGTGA	TGAACGCACA	CTGAAAAACA	CAGATATCAA	AGTTGTTGAC
9401	TTTGGAAGTG	CAACGTATGA	TGATGAACAT	CACAGTACTT	TGGTGTCTAC
9451	CCGGCACTAC	AGAGCTCCCG	AGGTCATTTT	GGGTCAGTAG	ACACCAGGCT
9501	TTCTAATATT	ATAATTGAAG	AAGAGATTTT	TGTTCTTTAC	AGCTTTACTG
9551	GTGGGGTGAG	GAAGTATGAT	CTTCTCAGCA	GGATTCAGAA	AACGTTTTCT
9601	ATTTTCTAAA	AAAATGTGTG	GACATTGCTA	TAAATACTTT	TCCTGAGTGG
9651	TAAACATGTG	ATACTGTCTG	GGAAAGATAT	TCCAGGTGGT	GGTTATTTTT

FIGURE 3C

9701	GAACAAGTAA	ATCTTAAATG	ATCATAAGAG	AACAGGCTGT	GTTAGCTAAA
9751	TGCATCAAAAG	AAATGTGATT	TTGAAAGTTAT	ATGAGTACCT	ATTTTCATGC
9801	CATCACAAAA	GCACATGGCT	GGTAAAAATA	CTGAGGAAAC	TGGTTGGCAG
9851	ATGTCTAGAA	TATAGGATGG	ATAAAGGTCA	AGAGAAGAAA	GAGGCTTCTC
9901	TAAGAGCTCC	TGTGATAACC	CTTGATGTGA	GAAAGTCTGG	GAAAGAAAAAT
9951	GAGTTAAGGT	GCAGAGTTTT	CAAATAAGAA	GGGACTTATT	AAGGGAGTGT
10001	TATGCCTCAA	CATTA AAAAGT	TATAGATCAG	GTGTGTTAAT	AAATCAGGGA
10051	AGTCAGAGAT	TGGCTTGGGA	GCTTGGAGAC	ATTGGGAAAC	ATTCAGATCA
10101	GGCATATCAA	GAGAGTTGAA	TGTAATAAGC	TGATTACTTA	GCCTAAAGTT
10151	AGGTCCAAC	GAGGTTAGAT	TGTAAAGCAT	TTTTGTGGAA	TCGTATTTTA
10201	ATACTTTTTA	CTTTTTTTGT	TGAACAACG	GGACTTGGTA	GTTCAGAATA
10251	GGAGTGTA	AGCAAACTCT	TGATACTTAC	CTAGAGTAGA	GTAGTAAAGG
10301	AGTGAGGAAA	TCAAGAATCC	TGTGCAGCTC	TTGCCACAG	AACTTCCCTT
10351	GATGACAGAA	ATGTTCCATT	TCTGCACTGT	CCCATATGGT	AGCCACTAGT
10401	CACTGTGCGT	GACTGACTAC	CTTGTAGTGG	GGCCAGTGTG	ACTGAGGAGA
10451	ACTGAGTTTT	GAATTTACAT	TAATTTTATT	TCAGATTTAA	ACAGCCACAT
10501	GTGGCTAGTG	GTTACCATAT	TGAACAAGCA	CAACTCTTAG	AGCTTGTCTT
10551	TTAAATGCGT	AATAATAGGG	TTTCTGCGTA	GTACAAATTG	AAAGGAGCTA
10601	CTGTGTAAGG	GTA AAAAGAAA	GCAATATGGG	AAGAGATAGT	GGACAGAGAG
10651	GTATTTTTCAG	AGATTAGAAAG	GCAATAGATT	CCTCATTTT	AGAATCAGAT
10701	TTTTCCCCAA	ATATTTGGCA	TTTTTTCTTT	GTTATTGGTA	TATCAAACAG
10751	TGGTGTCATCG	TACAGTGTGC	TATCCTAGAT	TGAGTAAAT	ATAGTATATA
10801	GTAACCCCCC	CCTTTTTTTT	TTCTTTGAGA	TGGAGTTTCA	CTTTGTCACC
10851	CAGGCTGGAG	TGCAGTGGTA	CCATCTCGGC	TCACTGCAAC	CTCCACCTCC
10901	CAGGTTGCGC	CGATTCTCCT	AACTCAGCCT	CCTGAGTAGC	TGGGATTACA
10951	GGTGCCCAACC	ACCACACCCG	GCTAATTTTT	ATAGTTTTTA	GTAGAGATGG
11001	GGTTTCACCA	TGTTAGCCAG	GCTGGTCTCG	AACTCCTGAC	CTCAGGTGAT
11051	CCTCCTGCCT	CGGCCTCCCA	AAGTGCTTGG	ATTACAGGCG	TGAGCCACCG
11101	CGCCCCGCCA	AGGATTTTTT	TTTTTTAATT	TTTATGTTTT	TTATAACAGA
11151	GACAGGGCCT	CACCATGTTG	CACAGGCTGG	TCTCGAACTC	CTGGGCTTAA
11201	GTGATCCGCC	TGCCTTGGCC	TCCCAAAGTG	CTGGGATTAT	AGGTGTGAGC
11251	CACCGCACCC	ACCAGAAATAT	GGTCAATCTT	ATTAATAAAG	TTCCAAATGT
11301	GGCCAAGCAA	GGGATAGTAC	AAATCTGAAA	TTGGAGTCCC	TGGCCTTGAG
11351	GAGAAAAGAT	CAGGAGATTG	GGAGAATAGA	AAGGTCCTTT	GTTTGTGGAG
11401	TGAGGATGAA	GGCATAATGC	AATTGGAGGG	GAAAATGTAG	TCAGGTGCTA
11451	GAGTTGAAGT	AGGCAGTTGG	CCTTATGTTG	GGTATAAAG	CTAACTCATC
11501	CAAGAATGAG	ATGATTTAGA	ATGGTGTACT	GCAGAAGATT	ACAGTCACCT
11551	GGGAAAAGAC	TAAATTGGGA	GATAGGAGTG	GTTGAAAAAT	AAAACTTTTT
11601	TTTTTTTTTTG	AGACGCACTC	TTGCACTGTC	ACCCGGGCTG	GACTGCAGTG
11651	GCACGATCTC	GGCTCACTGC	AACTTCTGCC	TCCTGGGTTT	AAGCGATTCT
11701	CCTGTGTCAG	CCTCCCAAGT	AGCTGGGCTT	ACAGGTGCC	GCCACCACGC
11751	CCAGCTAATT	TTTTGTATTT	TTAGTAGAGA	TGGGGTTTCA	CCACATTGGC
11801	CAGGCTGGTC	TCCAACCTCT	GACCTTGTGA	TTACCTGCC	TTGGCCTCCC
11851	AAAGTGCTGG	GATTACAGGT	GTGAGCCACC	GTGCCTGGTT	GAAAAATAAA
11901	ACTTTTATGA	GGTCCAAGCT	CTAGCATTTA	CGGATTTTGT	ATGTGTTAAT
11951	AGGTAGAAAC	CATGCTCCAT	TATTTATTTA	TTTATTTTTT	GAGACAGAGT
12001	CTCACTCTGT	TGCCTGGCCT	GGAGTGCAGT	GGTGCAATCT	CAGCTCACTG
12051	CAACCTCTGC	CTCCCGGGTT	CAAGCGATT	TCCTGCCTCA	GCCTCCTGAG
12101	TAGCTGGGAT	TACAAGTGCA	CACCACCACA	CCCAACTAAT	TTATATATAT
12151	ATATATATAT	ATATTTTTAAA	ATTTTTATTT	TTTATTTTTG	TTATTTGTTT
12201	ATTTATTTTT	TTGAGATGGA	GTTTTGCTTT	TATTGCCAG	GCTAGAGTGC
12251	AGTGGCGCAA	TCTCAGCTTA	CTGCAACCTC	TGCCTTCCGG	TTTCAAGCCA
12301	TTCTCCTGCC	TCAGCCTCCC	AAGTCACTGG	GATTACAGGC	GTCTGCCACC
12351	ACGCCCAGCT	AATTTTTTTTG	TATTTTTAGT	AGAGACGGGG	TTTCACCATG
12401	TTGGTCAGAC	TGGTCTCGAA	CTGCCAACCT	GGTGATCCAC	CCGCCTCGGC
12451	CTCCCAAAGT	GCTGGGATTA	CAGGCATGAG	CCACCGCGCC	TGGCCCATGC
12501	TCTATTATTA	TCCATTTGTT	CAAATGACAG	ACACTGGAGC	GGATGGTTAA
12551	CAAAAATGAC	TTAAGTCATT	ATATATTGAC	TTGAATATAT	TTCTTCTTTT
12601	ATCTTTAACT	TCAGTGATAA	TGAAAGTAAT	TGAAATGTCT	TTGAATGTAG
12651	ATTTTATTTA	TACATTTTTT	AACTAAATAT	TTGATCTTTG	AAATATTTAA
12701	ATATCTATGT	GGTTGGTTCT	TTCTCCTTCC	CAGTCAGTAT	AGATTTAAGA
12751	AGGCTAGATG	TTTTATTCTG	ATCTGAATAA	TACTGTCATT	GAGAATTCCTG
12801	AAGGAGAAAT	TATATAAAAT	CATGTATAGA	CAGCGCCGAT	GTTTATGTAT
12851	AGATCCCTCT	CTGAGCTCCA	ATGTGTCTGT	AATTTCTGCT	TATAGGTGAA
12901	ACTGCTTAAA	ATTCCCATTA	TACCTTTTAT	ACAATTTGTG	CAAAACGGTA

FIGURE 3D

12951	ATATTTCTCT	TAACGGAAGA	AGTAAACTCA	TGCATCAAGC	TGATGATAAT
13001	TGATAAGGCA	TTAGTAATTT	CATTCTGAGG	ATAATTATAA	ACCTGTATTT
13051	GTGCTAATAA	AATATAAAAA	TTCTTGGACT	AACCATGAAC	TGAGCATAAT
13101	AATGGTTTTA	ACAGCAGTGC	TCTCCCATTA	TATAAACAGT	TCAGAGACTA
13151	TGGAATATTT	GCACGAATTG	GTTGTATACT	TGGAAAATGG	TAGCCCCCTT
13201	TTATTTTACA	TAACATGCAC	CCCTCCCTAG	TTAGAATACT	GTGTCTTGAT
13251	GTGAGCATAT	GGACTATGGA	GTGTGTTGAA	TAGCATTTCG	TGTAAAACTA
13301	GAACATAAAA	CTCTGAATTT	GGTGTCTTAT	TCTCCCAAAT	GGGTTCTGTA
13351	AAGGGAGCAC	TCATATAGGG	AAGGATTTAA	TGTAATGTCA	ATTAATAAGTT
13401	TTTGACATAG	AAAATGTTTC	TATTTGTTTT	AAAATAGCTT	TAGGTTGGTC
13451	TCAGCCTTGT	GATGTTTGGA	GCATAGGTTG	CATTCTTATT	GAATATTACC
13501	TTGGTTTCAC	AGTCTTTCAG	GTACGTGGCT	AGTAAATTCC	ATTTAATAAT
13551	TCATAACAAA	TTGTAAACGT	TAAAGGTATG	CTAAAGTTTT	GACTTCCATA
13601	TTGGAAAATT	GCCATACATC	ATTATTCTTG	AGATTAAAAAC	TTAGGCCAAAA
13651	TGGTCATTCT	TTAAAAACCAC	AGTTGAATGA	AATATTACTA	TGAGTGAGTG
13701	ATCATAGTTA	ATTTTGCATG	TGATTAGTGT	TTGTAAACACA	TGGTTCATAT
13751	ATGGTTCATA	CTGTCTCCTT	TTTTAAATTG	TAGAGCTTCT	TCATAAATTT
13801	GCAGTAGTGT	TAATGTGGCC	AGTTTTTCAGT	TATAGTTATG	TTGACTATCA
13851	ATATGGCCAT	GAACGAGTCA	CTTATTCCTT	TTTATAAAAAG	AATTCAGGAA
13901	CAACAAGGGA	TTGTATTTTA	CTCTTAAGTA	TTAAGCATCT	ATAATGTCTT
13951	AGGCATTTCT	AAGTATAAGT	ACATAAAGGT	GAAGAGACAA	CATCTTTCTC
14001	AAGTCATGCA	AAAGACATTG	GAAAGTTATC	GCAGTATAGT	GTAGCATTTG
14051	CTGTGATGGA	ACAACGTAGA	AAGTGTAAGT	AGGGAGGGCC	AGGCGGGGTA
14101	GCTCACACCT	GTAATCCCAG	CACTTTGGGA	GGCTGAGGTG	GGTGGATCAT
14151	GAGGTCAGGA	GATCGAGACC	ATCCTGGCTA	ACATGGTGAA	ACCCTGTCTC
14201	TACTAAAAAGT	ATAAAAAATT	AGCTGGGCGT	GGTGGCGGGC	GCCTGTAGTC
14251	CCAGCTACTC	GGGAGGCTGA	GGCAGGAGAA	TGGCGTGAAC	CTGGGAGGCG
14301	GAGCTTGCAG	TGAGCGAGAT	CATGCCACTG	CACTCCAGCC	TGGACAACAG
14351	GGTGAGACTC	TGTTGCAAAA	AAAAAAAAAAAA	AAAAAAAAAAG	ACAAAGTGTA
14401	GGTAGGGAGA	ACCCAGGAAA	GGTTAATAAT	TACTTTAGAG	AAGGCGTCAC
14451	TGAGAACATA	GGAAGAGGAG	GAGGAGTTAG	AAAAGTGGAG	TGCAATGGGC
14501	ATATAAGGAA	GAAGAAATAG	TATCTGTAAG	TGCACAGAGG	AGTAAAGGAA
14551	CATATTCTAC	TCAGGGAAGA	ATAGCGTTGT	CAGAGTGTCT	TGTATAAATG
14601	GGAAAATTAT	AACAATAGGC	AAGGATCAAT	TCATAAAAGA	CTTCGCAAGG
14651	TATTGGTTTT	ATCCTAGAAG	TCAGTGGATT	CCAAAAGTAG	ACTGGTCCAA
14701	AATGAAAATG	GTTGTCTAGG	TTTGCCATTG	TGACCCTTAT	TTAGAGATTA
14751	TCCCTCCTGC	TTTTTTTTTT	TTTAATGTCT	CTTTTATGTA	ATGATAGTCA
14801	TAGTTGTTGG	TAGTTTGCTT	TTAAAAATAA	AAAGTCCTTA	ATTGGTAAAA
14851	CAAAAAGTAG	GAAACTCTAC	TTTCTTTTCC	ACTCTGTCCT	TAAGTTGTAC
14901	TTACATCTGA	AATCTTAATT	TTTTTTTTTT	TTTCCCTGAG	ATGGAGTCTC
14951	ACTGTGTGAC	CCAGGCTGGA	GTGCAAGTGC	GCAACGTCAG	CTCACTGCAA
15001	CCTCTGCCTC	CCGGGTTCAA	GTGATTCTCA	TGTCTCAGCC	TCCCAAGTAG
15051	CTGGGATTAC	AGGCACGAGC	CACTACACCC	CACTAATTTT	TTGTATTTTT
15101	AGTAGAGGGT	TTTGCTGTGT	TGACCAGGCT	GGTCTCGAAC	TCCTGACCTC
15151	AAGTGATCTA	CCCTCCTTGG	CCTCCCAAAG	TGCTGGGATT	ACAGGTGTGA
15201	GCCACCGCAC	CCAGCCTGAA	ATTTAAATTC	TTGAAAGCTT	TAGGTGATGC
15251	AACCAATTGAA	GAACTTTAAA	TAGGGTCATG	GTATGATCGA	GGTGTGTGTG
15301	TGTTTTGTTT	GGGGAAGAGG	GGCTGGAGAT	CCCAGCTAGT	ACTGTTGAGG
15351	TTGATTTGAA	GTTAGAGCAG	TGCAGGGGGC	ATGCAGCTAT	GATGGGCTAA
15401	GAGTCACTTA	GGCAGCTGTT	GCACAATGAT	GAATTCCTTG	TTCTGTTGGG
15451	ACCTCGCCAG	ATTTCTGTTT	CTGTCTAATC	TGTAGAGATC	CTGTTGAAAA
15501	GTAATCTGAG	TTTATAGATA	AGTTTGATGT	CTTAGAATCA	TGGTTATTAA
15551	TCAGTTCTGG	GAGGTATTGT	CTGGTTTTGC	AGTGGTGAGC	TGTAGGGTCA
15601	AGAAAAAGTT	AAGCAAAGTG	AATGCTTTCA	TCAATCTGAC	TAATATGAAA
15651	TGGATGCTTC	CGGTGATTTT	GTGATTATAA	ATCACTTTGA	GTTTTAAATG
15701	AAGTATATAT	TATTTGAGAG	GTGGTTTATA	TTTTAACTCC	ACCCTGCAAA
15751	ATACTCTTAA	ACTAAGGAAT	TTCTTTAAAA	TGTGAAGCTA	GTATTACTTA
15801	TTCTGTGATC	GTATCACAAC	GATTTGGAAG	CAATATGCAA	GGCACAGTAG
15851	TTGATAGATT	TCTTTTAAAA	TGTTTGATAA	CAGCCTCTGC	TCTCCAGAAC
15901	AAGGGTTAGC	AAACTTTGGC	CCATGGTGAA	ATCCTGCCTG	GTGCCTGTTT
15951	TTACAAAAAG	AAGAAGAGTA	TGCAATAGGG	ACCACTCATG	ACGAGCCAAG
16001	CCTAAAAATAT	TTACTATCTG	GCCCTTTACA	GAAGTTTGCC	AACCTCTGCT
16051	CTAGAAGCAT	ACCATTCCAG	CTGTAAGTTT	GACCGTTTTT	TGTATTCTAC
16101	TTAGGCCAAG	CCTCCGTTAC	TAATTTAAGG	ATATGTGCTT	TGACATGGGT
16151	TGATAGCTTA	ACTTTCTCTA	TATATGAGCT	ATATGACTTT	GAGGTAGTAT

FIGURE 3E

16201	CTTAACCTTT	TTGAAATTCA	TGTTCCCACA	TACCTAGCTC	AGAATTGTTT
16251	AGAGAATTAT	TGGGACTGTA	TGTATGCTG	TTGCCGTTGG	GTAAGTAAGTG
16301	TTAACAAGTG	AACTATTCAT	TGGGTAAGTG	ATGTTAATTT	TGGTTAAGCA
16351	GCTGATTAAA	TGAGGAGACA	GTTTTTCTGG	TAACCTTGCC	CAGTTATTCT
16401	TTAAACAGTG	TAAGAAGTGC	AAATAAAGAA	GGAAACTAAA	ATTTTAGATT
16451	AAACAAGTTA	ATGTGTTTGT	AGGGAAATGG	AGAGTACTAA	ATTTCTTTTT
16501	CTTACATGTT	TTAGACTCAT	GATAGTAAAG	AGCACCTGGC	AATGATGGAA
16551	CGAATATTAG	GACCCATACC	ACAACACATG	ATTCAGAAAA	CAAGGTATGT
16601	TTTAAGATTG	AAGACTTTTG	TTGGATATGT	GCAATAGCAT	ATATTCAAAC
16651	TACAGAAAA	CCAACGTTGT	TGTAATACTG	ATTCCAAGGA	CTATAGATT
16701	TGACTTTTTT	TTTTTTTTCT	GTACTGGAGG	TAACTTCTAA	CTTCATCTTA
16751	CTCCTTTTTT	TTTTTTTTGAG	ATGGAGTCTC	ACTCTGTCAC	CCAGGCTGGA
16801	GTGCAGTGGC	ACGATCTCAG	CTCACTGCAG	CCTCTGCCTC	CTGGGTTCAA
16851	GTGATTCTTC	TGCCTCAGCC	CCCTGAGTCG	CTGGGATTAC	AGGTGCCCCAC
16901	CACTATGCCT	GGCTAATTTT	TGTATTTTTA	GTAGAGATGG	GGTTTCACCG
16951	TGTTAGTCAG	GCTGGTCTTG	AACTCCTGAC	CTCAGGTGAT	CTGCCTGCCT
17001	TGGCCTCCCA	AAGTGCTGGA	ATTACAGGTG	TGAGTCACTG	CACTAGGCCA
17051	TGTTTTTAAA	AACTAATATA	ATAAAAAATA	TTTACCTTGT	GATCTAGTGC
17101	AGGGGTCCCC	AACCCCTCGG	AACTGGGCTG	TACAACAGGA	GGTGAGTGGC
17151	GGGTGAGTGA	GCATTATTGC	TGCCTGAGCT	GCACCTCCTG	TCAGATCAGC
17201	AGTGGCATTG	GATTCTCATA	GGAATGTGAA	CCCTATTGTG	AACTGCGCAC
17251	GTGAGGGGAT	TACGTTGCAT	GAAGGTTTCT	TATGAGAATC	TAATGCCTGA
17301	TGATCTGAGG	TGGAAGTTTG	ATTCCAAACC	ATCATCCCTC	CTCCCCGGAT
17351	CTGCTTCCAT	GAAACCGGTC	CCTGGTTCCA	AAAGGGTTGA	GGACCACTGA
17401	TCTAGTAAAC	AAAATGGCTT	TTGGGTTTTT	TTTGTTTTTT	TTTTTTTTTT
17451	AACTCAAGTT	TACGTTTGGC	ATAAGTGTTT	TCTTAGGCGA	TGTAAAAATA
17501	ATACATAGAA	TATGGAAAAA	CTTGTGTTTT	GGAATCATAT	CACTCTAAGT
17551	GTGAAATTTA	TTCTGTCTTT	AAACCAGCTG	ATATTCTTAG	ACAAGGTGGT
17601	ATTTCCAAAC	ACAGCTTCAT	CGCAGAAGCC	ACCGAGGGAG	TTCTTTAAAG
17651	ATTTCCAGCC	CCATTCTAGA	TCTAGTGAAA	ACAGAAATTT	AGGACTGGAT
17701	CCAGGGGGCC	CCTAGTTTTA	AGCTGACATT	GTTCCATATG	TGATAGGAAC
17751	AACTTAGTTG	AGAGACTAAA	ACCTCACAGG	GTGGAGGATA	TGAGGTGTCC
17801	GATATATAAT	TGTTGCTGAG	GTTTTTAAAA	ATTGTATGCA	TCTATATTAT
17851	ATAAGTCTAT	ACACTTAGAG	AGAGCTGCTT	TCCATGTCTC	CCCTCATGGG
17901	TGACAGGTAA	AGATACGACT	CTTGTTATTT	TACTAATCCA	GACTTTTTTT
17951	TTTTTTCTGT	AGAAAAACGCA	AGTATTTTCA	CCATAACCAG	CTAGATTGGG
18001	ATGAACACAG	TTCTGCTGGT	AGATATGTTA	GGAGACGCTG	CAAACCGTTG
18051	AAGGTAAAAA	AAAAAAGATT	AAAGGTTAAA	TAAACCACGT	GTTTGCACTA
18101	TTAATAATTT	TTTTTAAAC	AAAAACATTT	CTCCCCCAGG	AATTTATGCT
18151	TTGTATGAT	GAAGAACATG	AGAAACTGTT	TGACCTGGTT	CGAAGAATGT
18201	TAGAATATGA	TCCAACCTCA	AGAAATTACCT	TGGATGAAGC	ATTGCAGCAT
18251	CCTTTCTTTG	ACTTATTAAA	AAAGAAATGA	AATGGGAATC	AGTGGTCTTA
18301	CTATATACTT	CTCTAGAAGA	GATTACTTAA	GACTGTGTCA	GTCAACTAAA
18351	CATTCTAATA	TTTTTGTAATA	CATTAAATTA	TTTTGTACAG	TTAAGTGTAA
18401	ATATTGTATG	TTTTGTATCA	ATAGCATAAT	TAACTTGTTA	AGCAAGTATG
18451	GTCTTGATAA	TGCATTAGAA	AAATTTAAAT	TAATTTTTCT	TTTTGAAAT
18501	ACCATTTTTA	AAATACCTTT	AAATATCCTT	TGTGTCCAGT	GATAAATGTG
18551	ATTGATCTTG	CCTTTTGTAC	ATGGAGGTCA	CCTCTGAAGT	GATTTTTTTT
18601	GAGTAAAAGG	AAATCTTGAC	TACTTTATAT	TCTTAAAGGA	ATATTCTTTA
18651	TATACTTCAA	ATTTAGAACT	TAACTTTAAA	AGTTTTTCTT	CTGTAATTGT
18701	TGAACGGGTG	ATTATTATTA	ACTCTAGATA	AGCAGGTACT	AGAAACCAAA
18751	ACTCAGAAAA	TGTTTACTGT	TAGAAATTCTA	TTAAATTTTA	AGTGTTGTAT
18801	TCTTTTTTCAT	TGGGTGATGT	CAGGGTGATA	ACCAGACATT	CATGGAAAGG
18851	CATGCAGTTT	GTCCATTGTG	ACAGTTTGTG	TAATAAAACC	ACATACACAC
18901	TTTATTTAAG	ATTAAAACT	AACTGGAAAG	TCAGCTTGGA	AAATGGACAT
18951	TTCCAAGTAT	GTTTGGTGAG	TCACAGATAT	AAAAATAGAA	ATTCTGATGA
19001	GAGGTTTCAG	TTTTTAATAC	CAAGTCCTTA	GGAGTCTTAA	CATTGGCCAG
19051	CATCTGTTTA	TCAAAATGACA	TAAATACGTA	AACCTATAAG	AATTAAGTTT
19101	ATTAATTAGG	CAATTTATGT	CTGTGATAAT	TCTTACGGGA	GAAAGAGGAT
19151	TTGATTGGAA	AGCAGTTTGG	GAAGAAAGTG	CTGCTGAAAT	TTCCAGAATT
19201	TAATTGATTG	GTTACATAAA	CTTTTTGACT	TCAGCGTTTG	TTGTTGTTGT
19251	TCTTTTACTG	TCCTTGTTTT	CACATAAAAA	CTATATGGAG	CCAGGCACAG
19301	TGGCTCACGC	CTGTAATCCC	AGCATTTTGG	GAGACCGAGG	CAGGCGGATC
19351	ACCTGAGGCC	AGGAGTTTGA	GACCAGCTTT	GCCAACATGG	TGAAACCTTG
19401	TCTCTACTAA	AGATAACCAA	AAAGTGCTGG	GTGTGGTGGC	GGGCGCTGT

FIGURE 3F

```

19451 AATCCCAGCT ACTCTGGAGG CTGAGGCATG AGAATTGCTT GAATCCAGGA
19501 GGCGGAGTTT GCAGTGAGCT GAGATTGTGC CACTGCACTC CAGCCTGGGC
19551 GACAGAGCGA GACTCCGTCT CAAAAAAGA AAAAAACAAA CAAAAACAAA
19601 ACCCGGTATG TGGTAAATTA CTTAATTGGG CAAAAGAAAA AAATGTCTGT
19651 TGCTATGGTT CAGTCAGCCA GGTAGGAATA TTTTTTGTTG TAGAATTCCT
19701 AAGTGCTTAT TTCCAGATAC AGGTGAATTT TTGTTAAAAG TATCCCTGTT
19751 TCATAAGTGC ATTACACAAA TATTGGAGTT TTATCTGTTT AGGTTTTGTT
19801 TTTTTTTTAG ACTGAGTCTT GCTCTGTTGC CCAAGTTGGA GTGCAGTGGC
19851 GTGATCTCGG CTCACAGCAA CCTTCTTCCT CCTGGGTTCA AGCGATTCTC
19901 TTGACTCAGT TTCCCAGAGT GCTGGGATTA CAGGCATGTG CCACCAGGTC
19951 CTGCTAATTT TTGTATTTTT AGCAGAGGCA GGGTTTCACC ATGTTGTCTGA
20001 GGCTGGTCTC AAACCTCTGA CCTCAAGTGA TCTTCCTGCC TCGGCCTCCC
20051 AAAGTGCTGG GATTAAAGGC ATGAGCCACT ATGCCTGGCT AATCTGTTTA
20101 TGTATTTTAA ACATAAAATG CATGGGATTT TCTTGTAAGG CAAATAATGA
20151 AACCAAGCTT GGTTTTCTAT GTTACTTAGG GGCAACATTT GTCAATACAG
20201 TAAGGCTGTG TTCCTAAAGT AGACTAGGAG TTTTAAAGAA AGCTGAAACA
20251 AAAAGTTTAT TGTAGAATGA CTGCATACAT TATGTTTAGG CCTCTGATAT
20301 AGTCCAAATA CAGTGACTTT ATTTTCAGAAT AGTTGAACTG TATGTGATAA
20351 TTTTTTTTAA GAAGCATTTG ATGTTTAAAA ACAAGGTTTT TCCTGAGTTT
20401 ACCAGTGTAG CCCTACAGAT TAAGGTGTTT GCTATCCTTT ATTTTCCCCT
20451 TCATTTTATT TTTCCACTGC CATTGTACTA CCCAAGCCTC CTGTCCTTTC
20501 CCCCAATAAG TGCTTCAAGT TCCCAAATTA GTGTTTACTT TCTATGAAAA
20551 ACTCAGAGTA GCTGATCTCA GGATATAGGA GGAAAGAAAA ATATTCACAT
20601 TATTTCTTAC TAAGAAGTTA TTGATTGCTA ACCCCCTGTC TCTTCTGAAA
20651 ATTTACGTTT TTCACAAAGG GTATTTGCTA ATTTCTAGGC CTAATTCATG
20701 GAATTTTCGGG AATTAAAACG AAACCTTAAA AAATTAGGAT AGATGCAATG
20751 CTTAGAGGTT AGGGCAGTAC CTCTGGGATC ATTGAGTGTC TTTTGTCAAC
20801 CTTCTTCCC CTCTTCTTTG AGCTTTCAAG TTCCTACTCT TAATTGCCTT
20851 TTTTCCTTGT ATTTCTGAAC TCATTTTGTC AAGTTCCAAG GTTTTTTTTT
20901 TTTTTTTTTT TTTTGACAGT GCCTTGAGCT TCAACACTAA AAGGGAAAAA
20951 GATTTAGAAT GGCCAATGCA CATGAATCCT TTGTAATTTA GGTATTTTTC
21001 TTAATAATTT GATACCTCAT AGAATTACTA TTTCTAGAAA TTCCATTGAA
21051 TTGTTTCTAG AAATTCCATT GAAGTCAAGC TTGATTTTTT TAGGAGGCAT
21101 TTGTAAAGTG CAGCTAAGTA GATTATTTCC AGCTTGCTGC TGCTGCTCAT
21151 TTTCTTGAGG TTTTTTTTCA TCCATGCATT CATGAAAATT TTCAGAGTAG
21201 TTGAATTCAA TTGACTCCTG CTGACAGCAA GGGG
(SEQ ID NO: 3)

```

FEATURES:

```

Start: 2007
Exon: 2007-2059
Intron: 2060-3118
Exon: 3119-3341
Intron: 3342-4462
Exon: 4463-4553
Intron: 4554-4948
Exon: 4949-5015
Intron: 5016-8054
Exon: 8055-8171
Intron: 8172-8258
Exon: 8259-8425
Intron: 8426-9007
Exon: 9008-9102
Intron: 9103-9352
Exon: 9353-9482
Intron: 9483-13437
Exon: 13438-13520
Intron: 13521-16514
Exon: 16515-16594
Intron: 16595-17962
Exon: 17963-18053
Intron: 18054-18139
Exon: 18140-18277
Stop: 18278

```

FIGURE 3G

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
76	A	-	Beyond ORF(5')			
7980	C	T	Intron			
8571	C	T	Intron			
11257	T	C A	Intron			
11684	C	T	Intron			
13312	T	C	Intron			
17110	T	C	Intron			
17451	C	A	Intron			
20766	G	A	Beyond ORF(3')			
20914	T	-	Beyond ORF(3')			

Context:

DNA
Position

76 GCAGAAAAGTATAAAGATGGTAATCTCTGTAGGAAATTAGTCCCCATTATTTAGCTGTAA
AATTATAATTAATAA
[A,-]
AAAAATCTTTGTTTCTAAATCTTTGCCACTGATTATTTCTGAAAAACACTCCAGGAAG
AAGCATTTTTAAGTTAAAGCATGTGAACCTTTATTTCTTGCTACAGGTTTCATATTTCTTT
TTCTAGAGAGTTTGCCAAATTATACAACGTGCTCCTTCATGCTCTCACCAATCTTGGCTG
TTTTGAAAGGCCAAGAATAATGTTTTGATTAAACTGAATTTTTAAATTTCTAACGAATTT
GTCCGCTGTCATATATTTATTGATCATTTGAACATCTTTTTATTCTTAGCCTATTTATTA

7980 TAATATCTGTGCATTGCCTGGCACAGAGTAGGCCTAGCCTGGTAAATGAATGAATGCTTT
CAACAGTAGCATATCCTATTTTTGGTTTACATTTGTATATATCTTTTAAACTGTTGTTG
TATAAAATGTAATTAATTTTAAATTTCTAGGAGCAAACGTTAAACTCATAAGTATTAAG
GGAATTATCACTTCATATAAAGTATTTTATCAAAATGTTTTAAGAAGATGTTATATGGAA
TCTGCTATAATATGTTCTGAAAGATTATTTTAAATGGCATAGAGGAATTGGTAATTAAGA
[C,T]
TATGCTTTAGAGCATAACATGGCTTCAGCTCACTCTTGACATTTATCATTTTTATCTTA
ATTTTATTTTTAAGGGATGGCATGCATGTAGCAGTGAAAAATCGTAAAAAATGTAGGCCGT
TACCGTGAAGCAGCTCGTTCAAGAAATCCAAGTATTAGAGCACTTAAATAGTACTGATCCC
AATAGTGTCTTGTAAGTATAACTTTACCTAGGAGCCATCATATTACATGAAATATTCAG
GTTTCCATAAACTGAATTATTATTTTGCTCTGTTTTAGCCGATGTGTCCAGATGCTAGAA

8571 GATGCTAGAATGGTTTGATCATCATGGTCATGTTTGTATTGTGTTTGAACACTGCGGACT
TAGTACTTACGATTTTCATTAAAGAAAAACAGCTTTCTGCCATTTCAAATTGACCACATCAG
GCAGATGGCGTATCAGATCTGCCAGTCAATAAATTGTAAGTACACTTGATAAATCTTTAT
TTTTATTTATTTATTTATTTATTTATTTTGGAGACGGAGTCTCGCTCTGTACCCAGGCTG
GAGTGCAGTGGCGCTCTCGGGTCCCAGCAAGCTCAGCCTCCCGGGTTCACGCCATTTTCC
[C,T]
GCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCACCACCATGCCAGCTAATTTTT
TGTATTTTTAGTAGAGATGGGATTTTACAGTGTTAGCCAGGATGGTCTCGATCTCCTGAC
CTTGATTTGCCCCCTCGGCCTCCCAAAGTGCTGGGGTTATAGGCGTGAGCCACTGTGC
ACAGCAATAAATCTTTATTTTAAATATTTTTATGTTTGTACCTCCTTAACAATTAAGA
TAAATCTTTAAGCACCAGAAAACCTGTTTTTATTATACAAGCTATATATCAAATGTTGT

11257 CACCACCACACCCGGCTAATTTTTATAGTTTTTATAGAGATGGGGTTTACCATGTTAG
CCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGC
TTGGATTACAGGCGTGAGCCACCGCGCCCGGCCAAGGATTTTTTTTTTTTAAATTTTTATG
TTTTTTATAACAGAGACAGGGCCTCACCATGTTGCACAGGCTGGTCTCGAACTCCTGGGC
TTAAGTGATCCGCTGCTTGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCACCGC
[T,C,A]
CCCACCAGAATATGGTCAATCTTATTAATAAAGTTCCAAATGTGGCCAAGCAAGGGATAG
TACAAATCTGAAATTGGAGTCCCTGGCCTTGAGGAGAAAGAATCAGGAGATTGGGAGAAT

FIGURE 3H

AGAAAGGTCTTTGTTTGTGGAGTGAGGATGAAGGCATAATGCAATTGGAGGGGAAAAATG
TAGTCAGGTGCTAGAGTTGAAGTAGGCAGTTGGCCTTATGTTGGGTATAAAAGCTAACTC
ATCCAAGAATGAGATGATTTAGAATGGTGTACTGCAGAAGATTACAGTCACCTGGGAAAA

11684 GTCCTTTGTTTGTGGAGTGAGGATGAAGGCATAATGCAATTGGAGGGGAAAAATGTAGTCA
GGTGCTAGAGTTGAAGTAGGCAGTTGGCCTTATGTTGGGTATAAAAGCTAACTCATCCAA
GAATGAGATGATTTAGAATGGTGTACTGCAGAAGATTACAGTCACCTGGGAAAAAGACTAA
ATTGGGAGATAGGAGTGGTTGAAAAATAAACTTTTTTTTTTTTTTTTGGACGCGAGTCTTG
CACTGTCACCCGGGCTGGACTGCAGTGGCACGATCTCGGCTCACTGCAACTTCTGCCTCC
[C, T]
GGGTTCAAGCGATTCTCCTGTGTAGCCTCCCAAGTAGCTGGGCTTACAGGTGCCCGCCA
CCACGCCAGCTAATTTTTGTATTTTAGTAGAGATGGGGTTTACCACATTGGCCAGG
CTGGTCTCAACTCCTGACCTTGTGATTACCTGCCTTGGCCTCCCAAAGTGCTGGGATT
ACAGGTGTGAGCCACCGTGCCTGGTTGAAAAATAAACTTTTATGAGGTCCAAGCTCTAG
CATTTACGGATTTTGTATGTGTTAATAGGTAGAAACCATGCTCCATTATTTATTTATTTA

13312 TAGTAATTTTATTCTGAGGATAATTATAAACCTGTATTTGTGCTAATAAAATATAAAAT
TCTTGGACTAACCATGAAGTGAAGCATAATAATGGTTTTAACAGCAGTGCTCTCCCATTTAT
ATAAACAGTTGAGAGCTATGGAATATTTGCACGAATTGGTTGTATACTTGGAAAAATGGT
AGCCCCCTTTTATTTTACATAACATGCACCCCTCCCTAGTTAGAATACTGTGTCTTGATG
TGAGCATATGGACTATGGAGTGTGTTGAATAGCATTTGCTGTAAAACTAGAACTATAAAC
[T, C]
CTGAATTTGGTGTCTTATTCTCCCAAATGGGTTCTGTAAAGGGAGCACTCATATAGGGAA
GGATTTAATGTACTGTCAATTAAGTTTGTGATAGTAAAAATGTTTCTATTTGTTTTAA
AATAGCTTTAGGTTGGTCTCAGCCTTGTGATGTTGGAGCATAGGTTGCATTCTTATTGA
ATATTACCTTGGTTTACAGTCTTTTACAGGTACGTGGCTAGTAAATTCATTTAATAATTC
ATAACAAATTGTAAACGTTAAAGGTATGCTAAAGTTTGTACTCCATATTGGAAAAATTC

17110 CACGATCTCAGCTCACTGCAGCCTCTGCCTCCTGGGTTCAAGTGATTCTTCTGCCTCAGC
CCCCTGAGTCGCTGGGATTACAGGTGCCCACTATGCCTGGCTAATTTTTGTATTTTT
AGTAGAGATGGGGTTTACCGTGTTAGTCAGGCTGGTCTTGAAGTCTGACCTCAGGTGA
TCTGCCTGCCTTGGCCTCCCAAAGTGCTGGAAATTACAGGTGTGAGTCACTGCACTAGGCC
ATGTTTTTAAAACTAATATAATAAAAAATATTTACCTTGTGATCTAGTGCAGGGTCCC
[T, C]
AACCCCTCGGAAGTGGGCTGTACAACAGGAGGTGAGTGGCGGGTGAGTGAGCATTATTGC
TGCCTGAGCTGCACCTCCTGTGAGATCAGCAGTGGCATTAGATTCTCATAGGAATGTGAA
CCCTATTGTGAAGTGCACGTCAGGATCTACGTTGCATGAAGGTTCTTATGAGAATC
TAATGCCTGATGATCTGAGGTGGAAGTTTGATTCCAAACCATCATCCCTCCTCCCGGAT
CTGCTTCCATGAAACCGGTCCCTGGTTCCAAAAGGGTTGAGGACCACTGATCTAGTAAAC

17451 GGGTGAGTGAGCATTATTGCTGCCTGAGCTGCACCTCCTGTGAGATCAGCAGTGGCATT
GATTCTCATAGGAATGTGAACCTATTGTGAAGTGCACGTCAGGGATCTACGTTGCAT
GAAGGTTCTTATGAGAATCTAATGCCTGATGATCTGAGGTGGAAGTTTGATTCCAAACC
ATCATCCCTCCTCCCGGATCTGCTTCCATGAAACCGGTCCCTGGTTCCAAAAGGGTTGA
GGACCACTGATCTAGTAAACAAATGGCTTTTGGGTTTTTTTTTTTTTTTTTTTTTTTTT
[C, A]
ACTCAAGTTTACGTTTGGCATAAGTGTCTTCTAGGCGATGTAAAAATAATACATAGAAT
ATGGAAAAGCTTGTGTTTTGGAATCATATCACTCTAAGTGTGAAATTTATTCTGCTT
ACCAGCTGTATATTCTTAGACAAGGTGGTATTTCCAAACACAGCTTCATCGCAGAAGCCA
CCGAGGGAGTTCTTTAAAGATTTCCAGCCCCATTCTAGATCTAGTGAAAAACAGATTTTA
GGACTGGATCCAGGGGGCCCCCTAGTTTTAAGCTGACATTGTTCCATATGTGATAGGAACA

20766 ACTGCCATTGTAACCTACCAAGCCTCCTGTCTTTCCCCCAATAAGTGCTTCAAGTTCCCA
AATTAGTGTTTACTTTCTATGAAAACTCAGAGTAGCTGATCTCAGGATATAGGAGGAAA
GAAAAATATTACATTATTTCTTACTAAGAAGTTATTGATTGCTAACCCCTGTCTCTTC
TGAAAAATTTACGTTCTTCAAAAGGGTATTTGCTAATTTCTAGGCCTAATTCATGGAATT
TCGGGAATTAACGAACTTTAAAAAATTAGGATAGATGCAATGCTTAGAGGTTAGGGC
[G, A]
GTACCTCTGGGATCATTGAGTGTCTTTGTCAACCTTCTTCCCCTCTTCTTTGAGCTTT
CAAGTTCCTACTCTTAATTGCCTTTTTTCTTGTATTTCTGAAGTCAATTTGTCAAGTTC
CAAGGTTTTTTTTTTTTTTTTTTTTTTTTTGTACAGTGCCTTGAAGTCAACACTAAAAGGGA
AAAAGATTTAGAATGGCCAATGCACATGAATCCTTTGTAATTTAGGTATTTTCTTAATA
ATTTGATACCTCATAGAATTACTATTTCTAGAAATTCATTGAATTTGTTTCTAGAAATTC

20914 GAAGTTATTGATTGCTAACCCCTGTCTCTTCTGAAAATTTACGTTCTTCACAAAGGGTA

FIGURE 3I

TTTGCTAATTTCTAGGCCTAATTCATGGAATTTCTGGGAATTAACGAACTTTAAAAA
TTAGGATAGATGCAATGCTTAGAGGTTAGGGCAGTACCTCTGGGATCATTGAGTGTCTTT
TGTCAACCTTCCTTCCCCTCTTCTTTGAGCTTTCAAGTTCCTACTCTTAATTGCCTTTTT
TCCTTGATTTCTGAACTCATTTTGTCAAGTTCCAAGGTTTTTTTTTTTTTTTTTTTT
[T, -]
GACAGTGCCTTGAGCTTCAACACTAAAAGGGAAAAAGATTTAGAATGGCCAATGCACATG
AATCCTTTGTAATTTAGGTATTTTTCTTAATAATTTGATACCTCATAGAATTACTATTT
TAGAAATTCATTGAATTGTTTCTAGAAATTCATTGAAGTCAAGCTTGATTTTTTTAGG
AGGCATTTGTAAAGTGCAGCTAAGTAGATTATTTCCAGCTTGCTGCTGCTGCTCATTTTC
TTGAGGTTTTTTTTCATCCATGCATTGATGAAAAATTTTCAAGTAGTTGAATTCATTGA

Chromosome map:

Chromosome 5

FIGURE 3J